T2I-Adapter: Learning Adapters to Dig out More Controllable Ability for Text-to-Image Diffusion Models

Abstract

The incredible generative ability of large-scale text-to-image (T2I) models has demonstrated strong power of learning complex structures and meaningful semantics. However, relying solely on text prompts cannot fully take advantage of the knowledge learned by the model, especially when flexible and accurate controlling (e.g., structure and color) is needed. In this paper, we aim to ``dig out" the capabilities that T2I models have implicitly learned, and then explicitly use them to control the generation more granularly. Specifically, we propose to learn low-cost T2I-Adapters to align internal knowledge in T2I models with external control signals, while freezing the original large T2I models. In this way, we can train various adapters according to different conditions, achieving rich control and editing effects in the color and structure of the generation results. Further, the proposed T2I-Adapters have attractive properties of practical value, such as composability and generalization ability. Extensive experiments demonstrate that our T2I-Adapter has promising generation quality and a wide range of applications. Our code is available at https://github.com/TencentARC/T2I-Adapter.